

Deutsche Nationalbibliothek
30. Januar 2012

Änderungen im UTF-8-Zeichensatz

Sehr geehrte Damen und Herren,

die Deutsche Nationalbibliothek wird für Ihre Schnittstellen zur Datenbereitstellung den Konversionsdienst überarbeiten. Dies bedeutet, dass es Änderungen in der UTF-8-Kodierung unseren Datendiensten geben wird. Betroffen davon sind alle XML-Versionen der Datenformate und die Formate MARC 21 sowie GND-MARC 21 im UTF-8-Zeichensatz. Die Umstellung erfolgt zeitgleich mit der Einführung der GND (Gemeinsamen Normdatei) Mitte April 2012. Über den genauen Termin werden wir Sie noch informieren.

Konkret bedeutet das, dass die in UTF-8 erfassten Datensätze ohne weitere Konversionsprozesse im ursprünglichen Zeichensatz ausgeliefert werden. Die Kodierungsform wird nun einheitlich in der „zerlegten“ (decomposed) Form erfolgen, was bedeutet, dass Sonderzeichen zukünftig aus dem Grundbuchstaben und dem entsprechenden Diakritika bestehen. Umlaute werden z.B. als Grundbuchstabe und einem nachträglich auf den Vokal „combining diaresis“ (U+0308) ausgeliefert.

Die im August 2011 angebotene GND-Testdatei wurde entsprechend aktualisiert und liegt auf unserem Testserver <http://datendienst.d-nb.de/cgi-bin/mabit.pl?userID=testdat&pass=testdat&cmd=login> unter dem Namen GNDtestutf8.mrc zum Herunterladen bereit.

Die Datei sowie den Text dieses Rundschreibens finden Sie auch auf unserer Homepage unter http://www.d-nb.de/service/zd/rundschr_datendienst.htm

Falls sich Fragen ergeben, stehen wir Ihnen gerne zur Verfügung.

Mit freundlichen Grüßen

Heike Eichenauer

--

Heike Eichenauer
Deutsche Nationalbibliothek
Digitale Dienste
Adickesallee 1
D-60322 Frankfurt
Telefon: ++49 69 1525 1074
Telefax: ++49 69 1525 1636
<mailto:h.eichenauer@dnb.de>
<http://www.dnb.de>